CONSERVED STRUCTURAL NON-CODING RNA FAMILIES IN GAMMA-PROTEOBACTERIA

Stanislav V. Luban[1,2] & Daisuke Kihara[2,1]*
[1]Department of Computer Sciences, [2]Department of Biological Sciences
Purdue University, W. Lafayette, IN 47906
dkihara@purdue.edu

There has been a surprising recent discovery that RNA can actively regulate gene expression. Non-coding RNAs (ncRNAs), microRNAs, and RNA interference have drawn much attention these days. Detecting ncRNA genes in a genome sequence is not as easy as protein-coding genes, because non-coding RNA sequences have no easily exploitable statistical predispositions. However, conserved RNA secondary structures can be localized using probabilistic models of expected mutational patterns in pairwise sequence alignments. Here we have extracted ncRNAs from thirty microbial genomes using the QRNA program (S. Eddy *et al.*). First, we have prepared intergenic region sequences from the genomes. These were aligned pairwise against all of the other microbial intergenic regions using the BLAST program. QRNA scores and classifies the pairwise alignments of conserved intergenic regions as potential ncRNA loci by means of the pair stochastic context-free grammar model and as protein-coding regions or position-independent regions using the pair hidden Markov model. Then, the pairwise alignment candidates for ncRNAs were eliminated if the alignment did not exist bi-directionally (when query and subject entries are switched and the algorithm is re-run). The remaining alignments were clustered into families, and the families were trimmed and/or eliminated using existing ncRNA data and log odds confidence scores constraints. In total, 381 ncRNA families are identified from at least four distinct microbes. 18 families had contained candidate entries from at least 8 different organisms. In *E. coli,* more than 900 potential ncRNA regions were localized. To verify the results, identified ncRNAs were compared with known structural ncRNAs in the Rfam database and homologies were found. Conservation of the identified ncRNAs among gamma-proteobacteria is discussed. Future experimental verification of results in the wet lab is planned.

S. Luban was supported by Howard Hughes Summer Internship.